CHAPTER 03

DIAGNOSTIC ACCURACY OF DEVELOPMENTAL SCREENING IN PRIMARY CARE AT THE 18-MONTH HEALTH SUPERVISION VISIT: A CROSS-SECTIONAL STUDY

Meta van den Heuvel, Cornelia M. Borkhoff, Christine Koroshegyi, Weeda Zabih, Sijmen A. Reijneveld, Jonathon L. Maguire, Catherine S. Birken, Patricia C. Parkin

Published in CMAJ Open 2016;4(4):E634-640
Chapter 3
Diagnostic accuracy of developmental screening in primary care

ABSTRACT

Background: Communication delays are often the first presenting problem in infants with a range of developmental disabilities. Our objective was to assess the validity of the 18-month Nipissing District Developmental Screen (NDDS) with the Infant Toddler Checklist (ITC), a validated tool for detecting expressive language and other communication delays.

Methods: A cross-sectional design was used. Infants, 18-20 months, were recruited during scheduled health supervision visits. Parents completed both the 18-month NDDS and ITC. We assessed criterion validity (diagnostic test properties, overall agreement) for one or more “no” responses (1+NDDS flag) and two or more “no” responses (2+NDDS flag) using the ITC as a criterion measure.

Results: The study included 348 children (mean age 18.6 ± 0.7 months). The 1+NDDS flag had good sensitivity (94%; 95% CI 70%-100% and 86%; 95% CI 64%-96%), poor specificity (63%; 95% CI 58%-68% and 63%; 95% CI 58%-69%), and fair agreement (0.26) to identify expressive speech and other communication delays, respectively. The 2+NDDS flag had low to fair sensitivity (50%; 95% CI 26%-74% and 73%; 95% CI 50%-88%), good specificity (86%; 95% CI 82%-90% and 88%; 95% CI 84%-92%) and moderate agreement (0.45) to identify expressive speech and other communication delays, respectively.

Interpretation: The low specificity of the 1+NDDS flag may lead to over-diagnosis; the low sensitivity of the 2+NDDS flag may lead to under-diagnosis suggesting that infants who could benefit from early intervention may not be identified. The NDDS does not have adequate characteristics to accurately identify children with a range of communication delays.
INTRODUCTION

Developmental screening in early childhood is recommended by several leading professional organizations, including the Canadian Paediatric Society and the American Academy of Pediatrics.\(^1\) In contrast, the recent 2016 evidence-based guidelines from the Canadian Task Force on Preventive Health Care (CTFPHC) recommends against developmental screening using standardized tools.\(^4\) In the province of Ontario, an expert panel recommended an enhanced 18-month well-baby visit which was introduced in 2009.\(^1\) The panel recommended use of standardized tools including the Nipissing District Developmental Screen (NDDS), which is available free of charge and now commonly used. The conflicting recommendations from professional organizations, evidence-based guidelines, and policy makers present a dilemma for primary care practitioners. Several Canadian experts have expressed concerns regarding the CTFPHC recommendation against developmental screening, calling for more research on this topic.\(^5,6\)

Communication delays are often the first presenting problem in young children with a range of developmental disabilities including isolated expressive speech delay, autism spectrum disorders (ASD) and intellectual disability.\(^7-9\) Approximately 20% of toddlers have expressive speech delay, which resolves in about 50% by five years of age.\(^10\) Children with persistent communication delays have poorer outcomes in reading, spelling, math skills and increased emotional and behavioral difficulties.\(^10,11\)

The NDDS is a parent-completed questionnaire that addresses different areas of development including communication.\(^12\) The accuracy of the NDDS in identifying early communication delays has never been investigated. The Infant Toddler Checklist (ITC) was developed for early identification of children who have or are at-risk for developing a communication impairment.\(^7,13,14\) More recently, the ITC was assessed for the accuracy of detecting a range of developmental delays, including language delay, global developmental delay, and ASD.\(^15-17\) The ITC has been shown to be valid, reliable and has good diagnostic test properties.\(^7,14,16,17\) The ITC is freely available and is one of two measures recommended for early detection of ASD in Canada.\(^18,19\)

We speculate that many Canadian primary care practitioners will continue to use standardized developmental screening tools while awaiting further evidence to guide practice. We recently introduced the ITC screening tool into a network of primary care practices whose usual practice is use of the NDDS. Therefore, the primary objective of this study was to assess the validity of the 18-month NDDS compared with the ITC. As parents and clinicians commonly monitor the number of words spoken in early childhood as a measure of language development, our secondary objective was to describe parent responses regarding their 18-month old child from both screening tools.
Chapter 3

METHODS

Participants and Setting
A cross-sectional prospective design was used. Infants, 18-20 months, were recruited from January 2012 – February 2015 on a convenience basis, during scheduled health supervision visits from primary care practices participating in TARGet Kids!, a practice-based research network in Toronto, Canada (www.targetkids.ca). There are currently 9 large group practices in this network, each of which has between 3 and 10 practising physicians. Trained research assistants in the practices obtained survey and questionnaire data. The study protocol and cohort profile was recently published.

Children were included if their parents completed both the 18-month NDDS and the ITC. Participant characteristics were collected using a standardized instrument based on the Canadian Health Measures Survey. Exclusion criteria were: children with health conditions affecting growth, acute or chronic conditions (other than asthma), known severe developmental delay, and families unable to communicate in English. All data were entered into a web-based data management system (Medidata Rave ®). Research Ethics Boards approval was obtained from the Hospital for Sick Children and St. Michael’s Hospital, Toronto, Canada.

MEASURES

18-month Nipissing District Developmental Screen (NDDS)
The 18-month NDDS is a one page, 17-item parent-completed tool. Each question addresses one or more areas of development: fine motor, gross motor, emotional, social, self-help, learning, thinking, communication. Ten questions address communication. The NDDS is not a standardized tool. Recently, Cairney et al. evaluated the psychometric properties of the NDDS with the Bayley Scales of Infant Development (BSID, 3rd edition). In that study the NDDS had modest test-retest reliability (Spearman’s rho = 0.62) and low sensitivities ranging from 29% to 68% and specificities ranging from 58% to 88% to identify developmental delays.

In 2011, the wording of question six: “Does your child say 5 or more words?” was changed to “Does your child say 20 or more words?”. The NDDS authors provide a rationale for this change, highlighting the importance of early referral to a speech pathologist in children with a suspected delay and refer to the MacArthur-Bates Communicative Developmental Inventories (MB-CDI). We used the 2011 version of the NDDS.

The NDDS takes five minutes to complete. Response options are “yes” or “no”. One or more “no” responses (i.e., the child does not demonstrate the behaviour) indicates the need for further assessment and/or referral. This is known as the “one flag” rule (1+NDDS
flag). The “two flag” rule requires a minimum of two “no” responses for referral and/or follow-up (2+NDDS flag). Currently, the instructions of the 18-month NDDS recommend a “one flag” rule to follow-up with the healthcare and/or childcare professional regarding the child’s development. We assessed both the “one flag” and “two flag” rule for the NDDS.

**Infant Toddler Checklist (ITC)**

The ITC is a one page, 24-item parent-completed tool that can be used in primary care. The ITC was developed to measure language predictors in order to determine if a communication evaluation is needed. It is a standardized tool with normative scores for monthly intervals for children 6-24 months of age. The ITC was originally developed, normed and validated in two samples of US infants. In two further studies which included infants 12-24 months of age (n=232 and n=915), the ITC was found to have a sensitivity of 87% and 86%, and a specificity of 75% and 83%. In a large study in pediatric primary care in the US, the positive predictive value of the ITC was 75% for a range of developmental disorders (ASD, language delay and global development delay).

The ITC takes five minutes to complete and two minutes to score. Most questions are answered with a 3-point scale: “not yet”, “sometimes”, and “often”. Scoring produces three composite scores (social, expressive speech, symbolic) and a total score. The total score and each of the three composite scores are dichotomized as “concern/no concern” using the 10th percentile cut-point. It is recommended that a child should be carefully monitored if the expressive speech composite is below the 10th percentile (concern for expressive speech delay) and the ITC should be re-administered again in three months. Additionally, it is recommended that a child should be referred for an evaluation if the social composite, symbolic composite or the total score is below the 10th percentile (concern for other communication delay).

**STATISTICAL ANALYSIS**

Means, standard deviations and percentages were calculated to characterize the study participants. If a participant’s response to an individual question on the ITC was missing, we replaced it with a response corresponding to the level below the mode for that ITC domain (e.g., If most answers were “often”, then we replaced the missing response with “sometimes”). We assessed criterion validity of the 18-month NDDS by calculating diagnostic test properties and overall agreement. For the purpose of our analysis, the ITC was considered a criterion measure. Diagnostic test properties (sensitivity, specificity, positive predictive value, negative predictive value) of the 18-month NDDS compared with the dichotomized scores of the ITC were calculated (with 95% confidence intervals). For most screening tools, a sensitivity of 80% and specificity of 90% are generally recommended.
For developmental screening tools, because of behavioral noncompliance a sensitivity between 70%-80% and a specificity of 80% has been suggested.\textsuperscript{27}

Cohen’s kappa coefficients were calculated as measure of overall agreement between the NDDS and ITC. Cohen’s kappa gives a quantitative assessment of how well two tools agree. Levels of agreement below 0.20, between 0.21-0.40, between 0.41-0.60, and above 0.61 are considered poor, fair, moderate, and good, respectively.\textsuperscript{28}

Finally, we described the number of words spoken by 18-month old children according to parent responses on a comparable single question on both the NDDS and ITC calculating proportions. All analyses were performed with IBM SPSS 20.

**RESULTS**

**Participant characteristics**

We included 348 children with a mean age of 18.6 months (+ 0.7 months) (Table 1). The NDDS and ITC were completed at the same visit, with no intervention between measures. On the NDDS, 138 (39.7%) children had 1+flag and 54 (15.5%) had 2+flags (Figure 1). On the ITC, 16 (4.6%) children were identified with an expressive speech delay (concern on the expressive speech composite); 22 (6.3%) children were identified with a communication delay needing referral (concern on the social composite, symbolic composite, or total score).

**Table 1** Participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n = 348</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (months) mean (± SD)</td>
<td>18.6</td>
<td>(0.7)</td>
</tr>
<tr>
<td>Sex, male</td>
<td>192</td>
<td>(55.2)</td>
</tr>
<tr>
<td>Maternal education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- College / University</td>
<td>328</td>
<td>(94.3)</td>
</tr>
<tr>
<td>Nipissing District Developmental Screen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1+ flag</td>
<td>138</td>
<td>(39.7)</td>
</tr>
<tr>
<td>2+ flags</td>
<td>54</td>
<td>(15.5)</td>
</tr>
<tr>
<td>Infant Toddler Checklist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive speech delay\textsuperscript{1}</td>
<td>16</td>
<td>(4.6)</td>
</tr>
<tr>
<td>Other communication delays\textsuperscript{2}</td>
<td>22</td>
<td>(6.3)</td>
</tr>
</tbody>
</table>

\textsuperscript{1} Need to monitor: expressive speech composite below the 10th percentile

\textsuperscript{2} Need for referral: social composite, symbolic composite or the total score below the 10th percentile
Diagnostic test properties of the 18-month NDDS compared with the ITC

To identify children with an expressive speech delay, the 1+NDDS flag was sensitive (94%, 95% CI 70%-100%) but not specific (63%, 95% CI 58%-68%). The 2+NDDS flag reduced the sensitivity (50%, 95% CI 26%-74%) but increased the specificity (86%, 95% CI 82%-90%). To identify children with other communication delays (concerns on the communication composite, symbolic composite or total score), the 1+NDDS flag was sensitive (86%, 95% CI 64%-96%) but not specific (63%, 95% CI 58%-69%). The 2+NDDS flag reduced the sensitivity (73%, 95% CI 50%-88%) but increased the specificity (88%, 95% CI 84%-92%) (Tables 2 and 3).

**Figure 1** Percentage of children and the total number of questions failed on the 18-month NDDS
Chapter 3

Table 2 Diagnostic test properties of the +1 NDDS flag compared with the ITC.

<table>
<thead>
<tr>
<th>1+ NDDS flag</th>
<th>ITC</th>
<th></th>
<th>ITC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Positive</td>
<td>15</td>
<td>123</td>
<td>19</td>
<td>119</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
<td>209</td>
<td>3</td>
<td>207</td>
</tr>
<tr>
<td>Sensitivity (%) (95% CI)</td>
<td>94 (70-100)</td>
<td>86 (64-96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%) (95% CI)</td>
<td>63 (58-68)</td>
<td>63 (58-69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV (%) (95% CI)</td>
<td>11 (6-17)</td>
<td>14 (9-21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPV (%) (95% CI)</td>
<td>99 (97-100)</td>
<td>99 (96-100)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 **Need to monitor:** expressive speech composite below the 10th percentile  
2 **Need for referral:** social composite, symbolic composite or the total score below the 10th percentile

Table 3 Diagnostic test properties of the +2 NDDS flag compared with the ITC.

<table>
<thead>
<tr>
<th>2+ NDDS flag</th>
<th>ITC</th>
<th></th>
<th>ITC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Positive</td>
<td>8</td>
<td>46</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td>Negative</td>
<td>8</td>
<td>286</td>
<td>6</td>
<td>288</td>
</tr>
<tr>
<td>Sensitivity (%) (95% CI)</td>
<td>50 (26-74)</td>
<td>73 (50-88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%) (95% CI)</td>
<td>86 (82-90)</td>
<td>88 (84-92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV (%) (95% CI)</td>
<td>15 (7-28)</td>
<td>30 (18-44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPV (%) (95% CI)</td>
<td>97 (95-99)</td>
<td>98 (95-99)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 **Need to monitor:** expressive speech composite below the 10th percentile  
2 **Need for referral:** social composite, symbolic composite or the total score below the 10th percentile

Overall agreement of the 18-month NDDS with the ITC

The 1+NDDS flag had a fair agreement (Cohen’s kappa 0.26) with the need for monitoring or referral for communication delays on the ITC (concerns on any composite or total score). The 2+NDDS flag had a moderate agreement (Cohen’s kappa 0.45) with the need for monitoring or referral for communication delays on the ITC (concerns on any composite or total score).

Description of parent responses regarding the number of words spoken

Question 6 of the 18-month NDDS “Does your child speak more than 20 words?” was the
question with the greatest number (n=85, 24.4%) of parents responding “no” (Figure 2).

Question 17 on the ITC “About how many different words does your child use meaningfully that you recognize?” identified five (1.3%) children who spoke no words at all, 21 (6.0%) who spoke 1-3 words, 81 (23.3%) who spoke 4-10 words, 134 (38.5%) who spoke 11-30 words and 107 (30.7%) who spoke more than 30 words.

**Figure 2** Percentage of parents responding with “no” and “yes” to the 18-month NDDS individual questions

**INTERPRETATION**

We investigated whether the 18-month NDDS, commonly used by Canadian primary care practitioners, can identify communication delays in early childhood in a primary care setting. Compared with the ITC, the 1+NDDS flag had good sensitivity, poor specificity, and fair agreement, to identify expressive speech delay and other communication delays. Thus, the 1+NDDS flag cut-point may result in a large number of false positives leading to over-diagnosis. The 2+NDDS flag had low to fair sensitivity, good specificity and moderate
agreement, to identify expressive speech delay and other communication delays. Thus, the 2+NDDS flag cut-point may result in a large number of false negatives leading to under-diagnosis. Taken together, these findings suggest that the NDDS, at either cut-point, may not have adequate characteristics to accurately identify children with a range of communication delays.

There is only one study published evaluating the psychometric properties of the NDDS. The authors concluded that the NDDS had poor agreement with the reference standard (BSID 3rd edition,) and recommended not to use the NDDS on its own for the identification of developmental delays. Our findings are consistent with this research. In contrast to the NDDS, the ITC has undergone substantial development and validation over the past two decades. The ITC was recently assessed in 137 US primary care pediatricians at the 12 month visit. Of 10,479 ITCs completed, 1,318 were positive (12.5%). The 10.9 % prevalence of a positive ITC in our study is similar to that of the study in US primary care. In their recent guideline, the CTFPHC recommended against screening using standardized tools in children (1-4 years) with no apparent signs of developmental delay and developmental concerns. However, this recommendation was based on low-quality evidence and studies evaluating the diagnostic properties of the ITC were not included. A systematic review of 11 studies of primary care practitioners who worked without screening tools demonstrated that practitioners achieved poor sensitivity (range 14-54%) in correctly identifying children having a developmental concern, suggesting that sensitive screening tools may be valuable.

In our study, almost a quarter of the 18-month old infants did not speak 20 words or more on the NDDS. Almost a third of the 18-month old infants spoke 10 words or less according to the ITC. Data from “Wordbank”, an open database about children’s vocabulary growth that archives data from the MB-CDI of more than 5000 children, demonstrates a very wide vocabulary range at 18 months; from 13.9 words (10th percentile) to 269 words (90th percentile). The BSID (3rd edition) indicates that fewer than eight words spoken for a 24-month-old is 1.33 standard deviations below the normative mean. Our data demonstrates that many more parents responded “no” to question 6 than any other question on the 18-month NDDS. This single question accounted for more than 60% of infants with a 1+NDDS flag. As parents commonly monitor their child’s vocabulary, parents should be informed about the broad vocabulary range at 18-months to reduce unnecessary concerns identified with the NDDS. Additionally, it is important for child healthcare practitioners to be aware of the fair agreement and low specificity of the 1+NDDS flag for identifying children with expressive speech delay as compared with the ITC.

This study also has some limitations. First, children’s communication was not assessed with an independent, standardized measure for comparison such as the MB-CDI or the BSID or clinical evaluation. Second, our study population included only parents who were able to
communicate in English, and therefore results may not generalizable to other populations. Finally, the NDDS is a general development screening tool whereas the ITC was initially developed to target communication delays, so it would be expected that some items (e.g., gross motor) on the NDDS would not correlate with the ITC. However, the majority of questions on the 18-month NDDS address communication and the ITC has a high positive predictive value for detecting global developmental delay, making the comparison of both tools justifiable. Additionally, the authors of the NDDS did not provide independent domains (e.g., “communication domain”) therefore a specific validation was not feasible.

CONCLUSION

Our results demonstrate that infants with communication delays will not be adequately identified with the 18-month NDDS. The low specificity of the 1+NDDS flag may lead to over-diagnosis and cause unnecessary concern for parents. The lower sensitivity of the 2+NDDS flag may lead to under-diagnosis suggesting that infants who could benefit from early intervention may not be identified. The NDDS does not have adequate characteristics to accurately identify children with a range of communication delays. With communication delays often being the first presenting problem in young children with a range of developmental disabilities, the ITC may be a more promising screening tool for assessing developmental delay in early childhood. A rigorous research agenda is needed to investigate the use of developmental screening tools in primary care.5
REFERENCES


5. LeBlanc JC. Take home message of task force report: NOT the strong recommendation against developmental screening, BUT the need for rigorous research and practice. *CMAJ*. 2016: June, 10.


